AD-A087 108 NATIONAL TELECOMMUNICATIONS/INFORMATION ADMINISTRATIO--ETC F/G 9/4
VOICE PERFORMANCE MEASUREMENTS.(U)
JUN 80 W J HARTMAN, L E PRATT DOT-FA79WAI-013
UNCLASSIFIED FAA-RD-80-71 NL

1 OF 1
AD A
087108

END
DATE
FILMED
9-80
DTIC

Report No. FAA-RD-80-71

VOICE PERFORMANCE MEASUREMENTS

W.J. Hartman
L.E. Pratt

U.S. Department of Commerce
National Telecommunications and Information Administration
Institute for Telecommunication Sciences
Boulder, Colorado 80303

June 1980
Final Report

Document is available to the U.S. public through
the National Technical Information Service,
Springfield, Virginia 22161.

Prepared for

U.S. DEPARTMENT OF TRANSPORTATION
FEDERAL AVIATION ADMINISTRATION
Systems Research & Development Service
Washington, D.C. 20590

NOTICE

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
18 FAA-RD-80-71 19	AD-A087108	

4. Title and Subtitle	5. Report Date
6 VOICE PERFORMANCE MEASUREMENTS.	11 Jun[redacted]80
	6. Performing Organization Code: NTIA/ITS
7. Author(s): 10 W. J. Hartman [redacted] L. E. Pratt	8. Performing Organization Report No.
9. Performing Organization Name and Address	10. Work Unit No. (TRAIS)
Institute for Telecommunication Sciences National Telecommunications & Information Administration U. S. Department of Commerce Boulder, Colorado 80303	11. Contract or Grant No.: 15 DOT-FA79WAI-013
12. Sponsoring Agency Name and Address	13. Type of Report and Period Covered: 9 Final rept.
U.S. Dept. of Transportation Federal Aviation Administration Systems Research and Development Service Washington, DC 20591	14. Sponsoring Agency Code: ARD 61C

15. Supplementary Notes

16. Abstract

Subjective intelligibility scores are determined for various combinations of signal-to-noise ratios, clipping, filtering, and pre-emphasis/de-emphasis. These are compared with scores derived using linear predictive coding (LPC). An overall correlation of 0.88 is obtained between the two sets of scores.

12 26

17. Key Words	18. Distribution Statement
Distortion, intelligibility, interference, objective scoring.	Document is available to the public through the National Technical Information Service, Springfield, Virginia 22151.

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	26	

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

410937

METRIC CONVERSION FACTORS

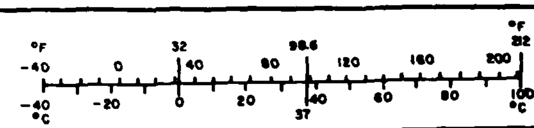
Approximate Conversions to Metric Measures

Symbol	When You Know	Multiply by	To Find	Symbol
		LENGTH		
in	inches	*2.5	centimeters	cm
ft	feet	30	centimeters	cm
yd	yards	0.9	meters	m
mi	miles	1.6	kilometers	km
		AREA		
in^2	square inches	6.5	square centimeters	cm^2
ft^2	square feet	0.09	square meters	m^2
yd^2	square yards	0.8	square meters	m^2
mi^2	square miles	2.6	square kilometers	km^2
	acres	0.4	hectares	ha
		MASS (weight)		
oz	ounces	28	grams	g
lb	pounds	0.45	kilograms	kg
	short tons (2000 lb)	0.9	tonnes	t
		VOLUME		
tsp	teaspoons	5	milliliters	ml
Tbsp	tablespoons	15	milliliters	ml
fl oz	fluid ounces	30	milliliters	ml
c	cups	0.24	liters	l
pt	pints	0.47	liters	l
qt	quarts	0.95	liters	l
gal	gallons	3.8	liters	l
ft^3	cubic feet	0.03	cubic meters	m^3
yd^3	cubic yards	0.76	cubic meters	m^3
		TEMPERATURE (exact)		
°F	Fahrenheit temperature	5/9 (after subtracting 32)	Celsius temperature	°C

*1 in = 2.54 (exactly). For other exact conversions and more detailed tables, see NBS Misc. Publ. 286, Units of Weights and Measures, Price $2.25, SD Catalog No. C13.10:286.

Approximate Conversions from Metric Measures

Symbol	When You Know	Multiply by	To Find	Symbol
		LENGTH		
mm	millimeters	0.04	inches	in
cm	centimeters	0.4	inches	in
m	meters	3.3	feet	ft
m	meters	1.1	yards	yd
km	kilometers	0.6	miles	mi
		AREA		
cm^2	square centimeters	0.16	square inches	in^2
m^2	square meters	1.2	square yards	yd^2
km^2	square kilometers	0.4	square miles	mi^2
ha	hectares (10,000 m^2)	2.5	acres	
		MASS (weight)		
g	grams	0.035	ounces	oz
kg	kilograms	2.2	pounds	lb
t	tonnes (1000 kg)	1.1	short tons	
		VOLUME		
ml	milliliters	0.03	fluid ounces	fl oz
l	liters	2.1	pints	pt
l	liters	1.06	quarts	qt
l	liters	0.26	gallons	gal
m^3	cubic meters	35	cubic feet	ft^3
m^3	cubic meters	1.3	cubic yards	yd^3
		TEMPERATURE (exact)		
°C	Celsius temperature	9/5 (then add 32)	Fahrenheit temperature	°F



FEDERAL AVIATION ADMINISTRATION
SYSTEMS RESEARCH AND DEVELOPMENT SERVICE
SPECTRUM MANAGEMENT STAFF

STATEMENT OF MISSION

The mission of the Spectrum Management Branch is to assist the Department of State, National Telecommunications and Information Administration, and the Federal Communications Commission in assuring the FAA's and the nation's aviation interests with sufficient protected electromagnetic telecommunications resources throughout the world and to provide for the safe conduct of aeronautical flight by fostering effective and efficient use of a natural resource - the electromagnetic radio frequency spectrum.

This objective is achieved through the following services:

- Planning and defending the acquisition and retention of sufficient radio frequency spectrum to support the aeronautical interests of the nation, at home and abroad, and spectrum standardization for the world's aviation community.

- Providing research, analysis, engineering, and evaluation in the development of spectrum related policy, planning, standards, criteria, measurement equipment, and measurement techniques.

- Conducting electromagnetic compatibility analyses to determine intra/intersystem viability and design parameters, to assure certification of adequate spectrum to support system operational use and projected growth patterns, to defend aeronautical services spectrum from encroachment by others, and to provide for the efficient use of the aeronautical spectrum.

- Developing automated frequency selection computer programs/routines to provide frequency planning, frequency assignment, and spectrum analysis capabilities in the spectrum supporting the National Airspace System.

- Providing spectrum management consultation, assistance, and guidance to all aviation interests, users, and providers of equipment and services, both national and international.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

VOICE PERFORMANCE MEASUREMENTS

W. J. Hartman and L. E. Pratt*

Subjective intelligibility scores are determined for various combinations of signal-to-noise ratios, clipping, filtering, and pre-emphasis/de-emphasis. These are compared with scores derived using linear predictive coding (LPC). An overall correlation of 0.88 is obtained between the two sets of scores.

Key words: Distortion, intelligibility, interference, objective scoring.

1. INTRODUCTION

An objective method of determining articulation score was developed by Gamauf and Hartman [1977]. The purpose of this paper is to compare the results of this scoring with subjective (listener panel) scores for a variety of conditions affecting the intelligibility. The original objective method reported in Gamauf and Hartman [1977] is modified in this study, the main modification being the development of a hardware word alignment device which is described in Appendix A.

Previous tests for which the objective scoring was obtained were done using rf modulation and demodulation, with the noise and distortion introduced at the rf frequency. The tests reported here were done entirely at baseband.

2. VOICE SCORING METHODS

In order to develop an objective intelligibility measure for corrupted speech, a comparison must be performed between the distorted speech and the original noise-free speech. A subjective intelligibility measure of the distorted speech must also be available in order to judge the quality of the objective measure being used. Both of these requirements are met by first making a noise-free master tape of preselected speech, and then sending it through voice communication channels to be tested, and making a recording of the speech at the channel output. The comparisons between subjective and objective scores is thus obtained from the same set of output data.

*The authors are with the U.S. Department of Commerce, National Telecommunications and Information Administration, Institute for Telecommunication Sciences, Boulder, CO 80303.

The pre-selected speech samples to be sent over a voice channel for intelligibility scoring are phonetically balanced (PB) groups of isolated words as opposed to complete sentences or nonsense syllables. These P B words were used because subjective scores have been shown to be repeatable, which is a necessary criterion for this study because the objective measure will be repeatable. Eight PB word groups, each containing fifty isolated words were selected as the test speech. The resulting scores are called articulation scores (AS).

An analog tape containing all eight word groups and using both male and female trained speakers was obtained from the Army Electronic Proving Ground Electromagnetic Environment Test Facility (EMETF) at Fort Huachuca, Arizona. From this tape, a master analog tape was made that would be sent over voice channels and later compared with the recorded output of the channel. In order to perform this comparison, the two tapes must be aligned, which means that synchronization information must be included on the master tape before being sent across the voice channel. Because the tapes are digitied at a 10 kHz rate, the alignment procedure would also have to work in a digital format. It was found that a shift of plus or minus 10 samples of a 256-sample analysis window caused the predictor coefficients to vary less than 0.1% in all cases. Therefore, the synchronization procedure to be used was required to align two segments of digitized speech to within 10 samples.

A synchronization procedure that was found to meet the required 10-sample variation specification made use of a binary pseudo noise (PN) sequence. A length 127 binary PN sequence, generated at a 635 Hz clock rate was sent through a phase-continuous frequency shift keying modem using the two frequencies 1.2 kHz and 2.2 kHz. This PN signal was then placed before each word and after the last word of all eight word groups thereby creating the master analog tape with alignment capabilities. In the previous work, Gamauf and Hartman (1977), the correlation between PN sequences from the master tape and the tape from the system output was done in a computer. For this study, a hardware correlation device described in Appendix A was used for the alignment.

3. TEST RESULTS

Eight different test conditions, involving pre-emphasis/de-emphasis, channel bandwidth filters, channel weighting, clipping and interference were set up in accordance with the block diagram of Figure 1. Within each test condition, four values of a parameter (e.g., the signal-to-noise ratio), were used to give a set of 32 measurements. Table 1 gives the test and parameter values for the 32 tests. Details of the test condition are given in Appendix B.

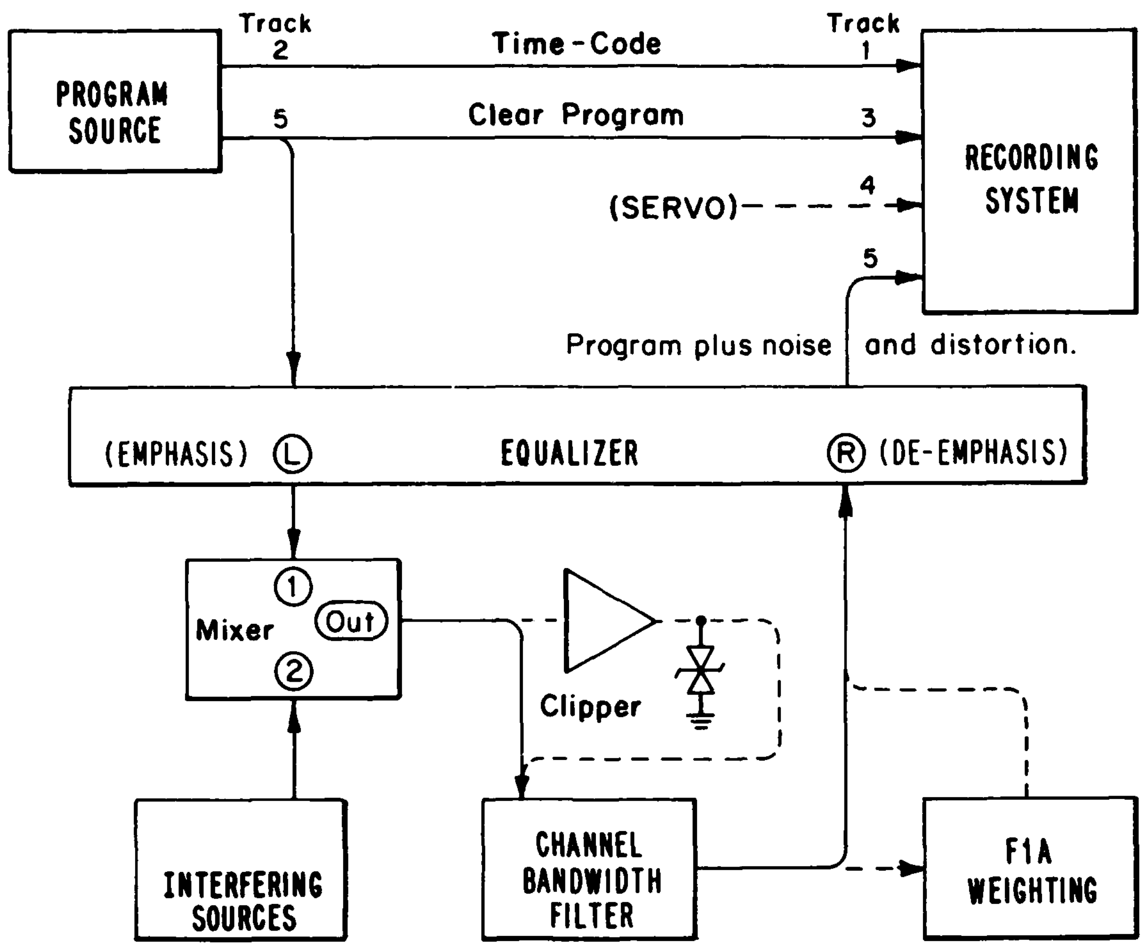


Figure 1. Block diagram of the test configurations.

Table 1. Test conditions.

Test setup #	Emphasis/ Deemphasis (dB)	Channel Bw	Channel Weighting	Interference Type	$\frac{S}{I+D}$	$\frac{S+I+D}{I+D}$	Condition	Word Group Number
1	0	W	flat	noise	-5	+1.4	1	361
					0	+3.0	2	312
					+5	+6.0	3	291
					+15	+15.0	4	265
2		W		clipp.		+7.7	1	275
	0	N				+12.9	2	305
	6	W				+14.9	3	214
		N				+17.8	4	283
3		W		noise	-5	+2	1	283
					0	+3.5	2	361
					+5	+6.7	3	312
	6				+15	+16.	4	291
4	3				-5	+1.5	1	265
					0	+3.2	2	275
					+5	+6.3	3	305
		W			+15	+15.7	4	214
5		N			-5	+2.0	1	214
					0	+3.5	2	283
					+5	+6.5	3	361
			flat		+15	+16.	4	312
6			F1A		-5	+3.0	1	291
					0	+4.5	2	265
					+5	+7.8	3	275
			F1A	noise	+15	+17.0	4	305
7			flat	2Khz	-15	0.2	1	305
				sine	-5	+1.2	2	214
				wave	0	+3.0	3	283
					+5	+6.2	4	361
8				Cross	-15	---	1	312
				talk	-5	---	2	291
					0	---	3	265
	3	N	flat		+5	---	4	275

The tapes were copied and sent to EMETF, Fort Huachuca, Arizona, for the subjective scoring.

For the objective scoring the output of the analog recorder was first conditioned as shown in Figure 2 and then fed to both the PN sequence detector and the digitizer as shown in Figure 3. The PN sequence detector printed the sample number corresponding to the corresponding digital sample whenever the PN signal on the tape aligned with a reference PN signal to allow accurate word alignment. The digitized signals were then processed according to the methods of Gamauf and Hartman to obtain the objective score.

Table 2 gives the subjective scores and objective scores for all the tests. No objective scores could be obtained for test 7 due to the combination of the analysis method used and the coherency of the sine wave. A modification of the method to resolve this problem is possible, but beyond the scope of the present effort.

The objective scores are plotted vs the subjective scores in Figure 4. The cross correlation coefficient between the two sets of scores is 0.88.

4. DISCUSSION OF RESULTS

The worst comparison between the objective and subjective scores appears to be for the clipping-narrow band case (test 2 conditions 2 and 4) which is not unexpected, and for the worst-case conditions that produced the lowest AS. It is mildly surprising however, that the scores agree so well for the case of cross-talk (test 8) since LPC (linear predictive coding) voice systems do not perform well with combinations of more than one talker at a time.

The results support the objective measure as a predictor for articulation score. Moreover, the cost reduction achieved using the hardware word alignment makes the process attractive for testing voice systems.

5. REFERENCE

Gamauf, K. J., and W. J. Hartman (1977), Objective Measurement of Voice Channel Intelligibility, FAA Report No. FAA-RD-77-153, October.

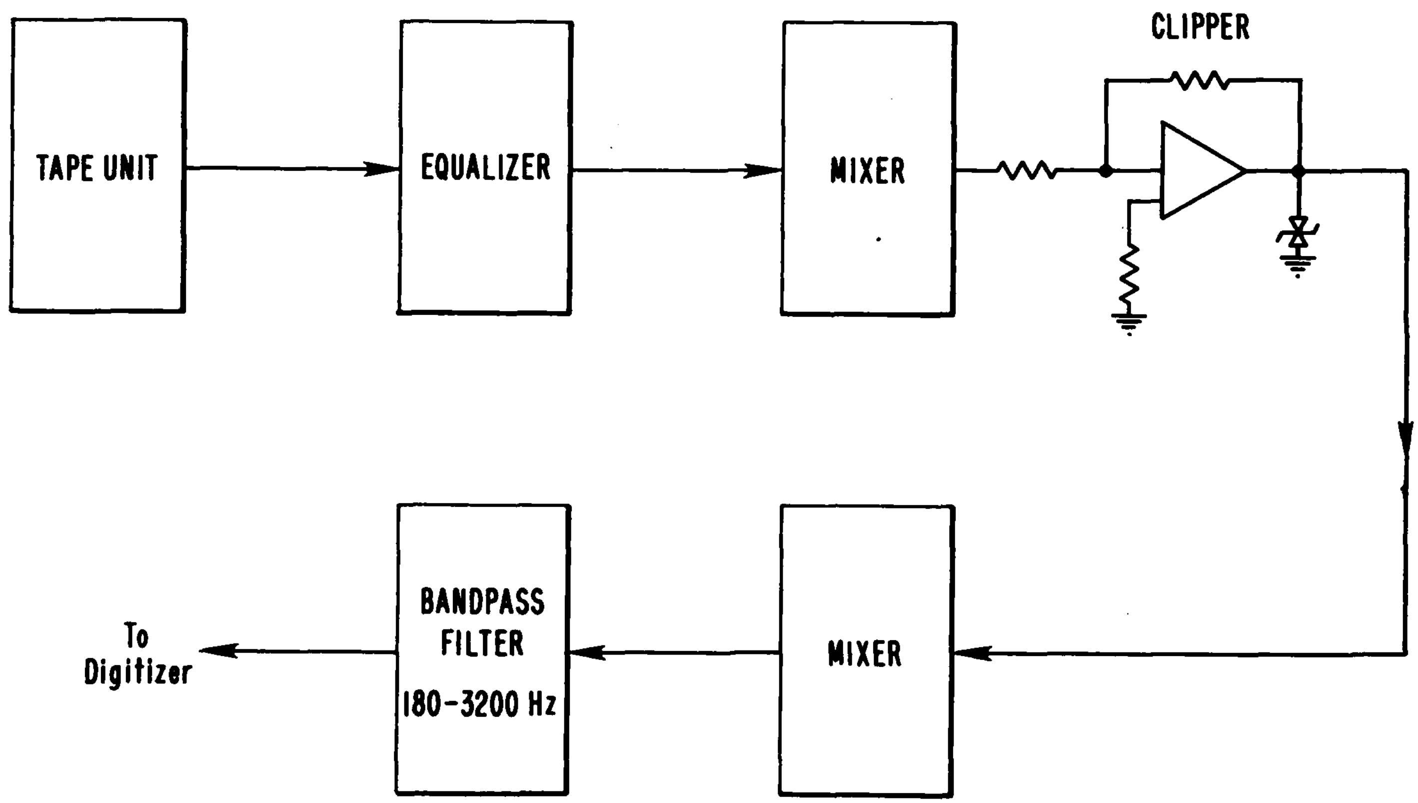


Figure 2. Block diagram of the signal conditioning network for digitizing the voice tapes.

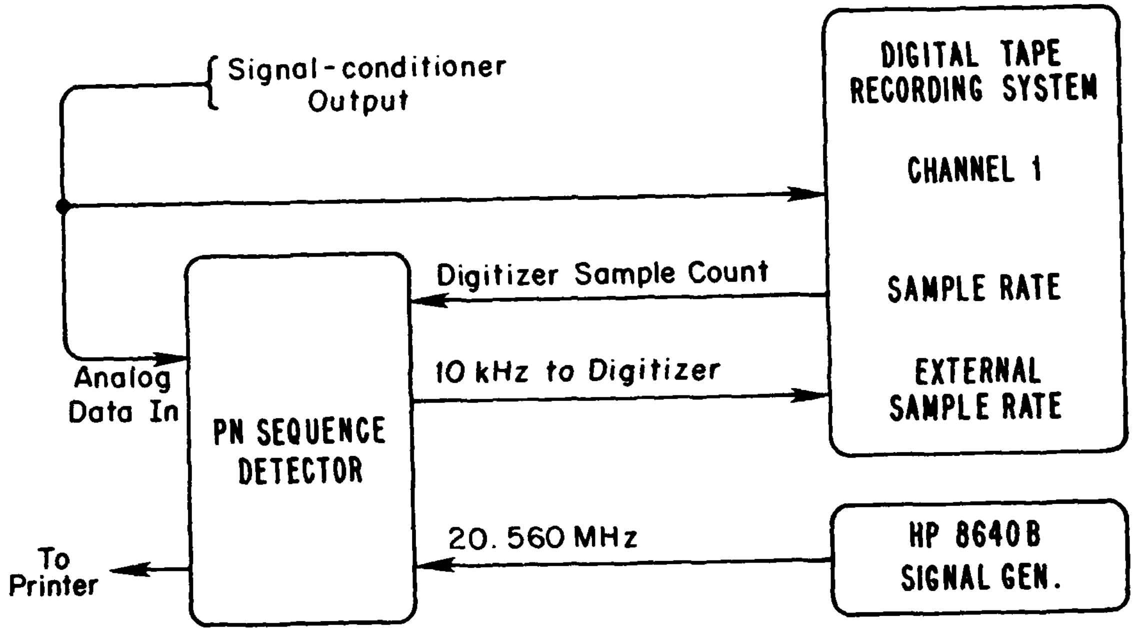


Figure 3. Block diagram showing the positioning of the PN sequence detector in the digitizing operations.

Table 2. Subjective and Objective Scores for the Tests Described in Table 1

Test #	Condition #	Objective Score/100	Subjective Score/100
1	1	.03	.28
	2	.48	.50
	3	.58	.65
	4	.83	.81
2	1	.53	.71
	2	.53	.69
	3	.91	.87
	4	.58	.85
3	1	.34	.16
	2	.47	.23
	3	.49	.55
	4	.78	.79
4	1	.44	.23
	2	.56	.44
	3	.63	.60
	4	.82	.88
5	1	.10	.14
	2	.20	.24
	3	.56	.44
	4	.67	.80
6	1	.15	.06
	2	.25	.12
	3	.32	.23
	4	.58	.62
7	1	---	.52
	2	---	.80
	3	---	.76
	4	---	.83
8	1	.04	.10
	2	.36	.46
	3	.68	.62
	4	.75	.71

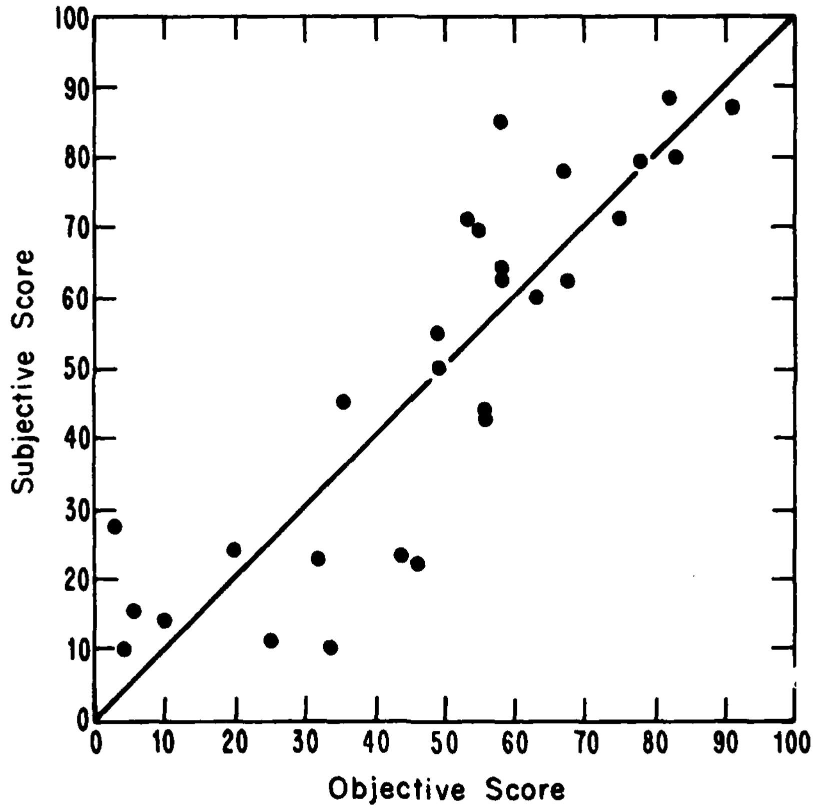


Figure 4. Comparison of the subjective (articulation score) and objective scores for 28 of the test conditions.

APPENDIX A: CORRELATION DETECTOR

The correlation detector was designed and built by ITS personnel. Its function is to find the point of highest correlation between an incoming signal and a stored reference signal. Use of a 10-KHz sample rate gave a 100 μs accuracy.

Refer to the block and schematic diagrams (Figures A1, A2, A3). The standard reference signal was digitized in 2048 samples, and only the polarity of the samples was permanently stored as the image in programmable read-only memories (PROM). The incoming analog signal is polarity detected and stored in random-access memory (RAM). The contents of the PROM and RAM are compared bit by bit in the exclusive- or gate-correlator and fed to the correlation counter. This is done 8 bits at a time in parallel because the PROM's and RAM's will not operate at the necessary clock rate of 20.56 MHz.

At the completion of each 2048 bit correlation, the output of the correlation counter is sent to a buffer (latch) and to a digital comparator for comparison with the current contents of the buffer. If the current count exceeds 1280 and is larger than the previous count, the current count is stored in the buffer and the digitized data sample number is stored in the sample counter. (This operation is shown on the block diagram only.) At this time a new signal value is brought into the RAM and the contents of the RAM are shifted by one to allow the correlator to operate on a new relative position of the data and the stored standard. The 2048th bit is lost because it has served its function and is no longer needed. This action also starts a 3-ms retriggerable time delay. This allows the circuit to keep looking for higher correlation values to insure that the highest value for the current word synchronization has been found. At the end of this 3-ms delay, a second 3-ms pulse is generated to cause the printer to print the digitized data sample number and to reset the buffer so the correlator can start looking for synchronization of the next word.

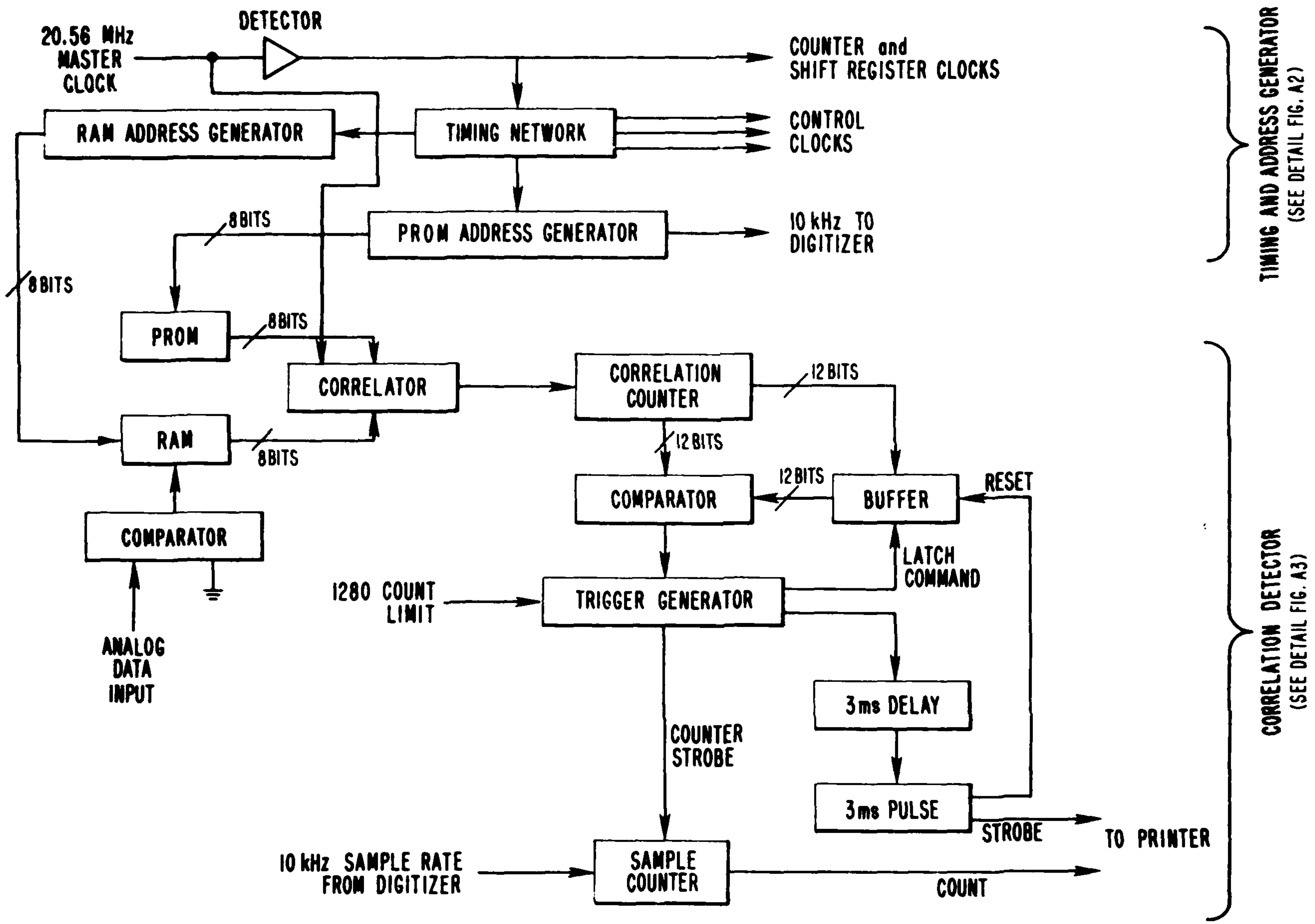


Figure A1. Correlation detector block diagram.

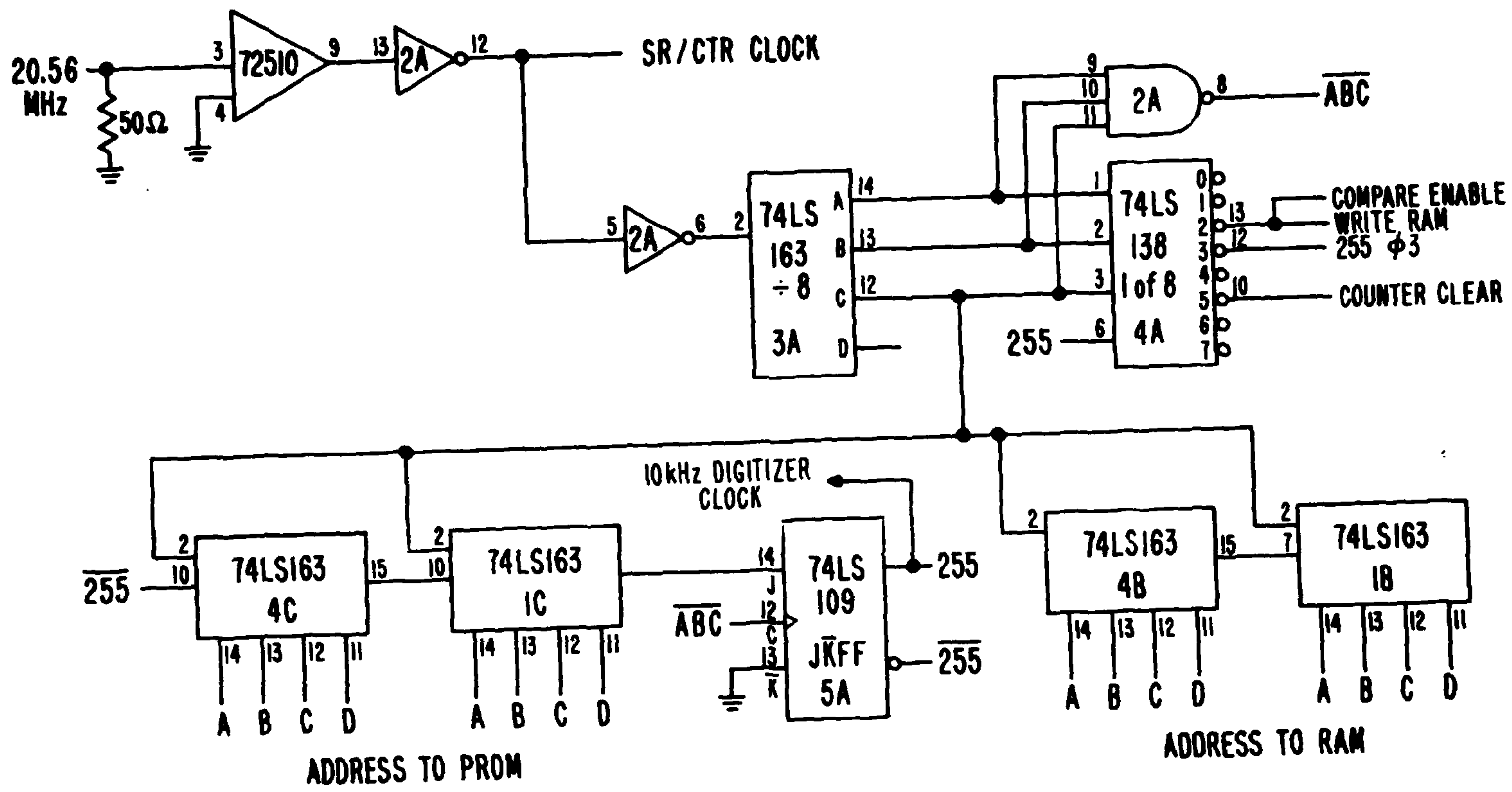


Figure A2. Timing and address generators.

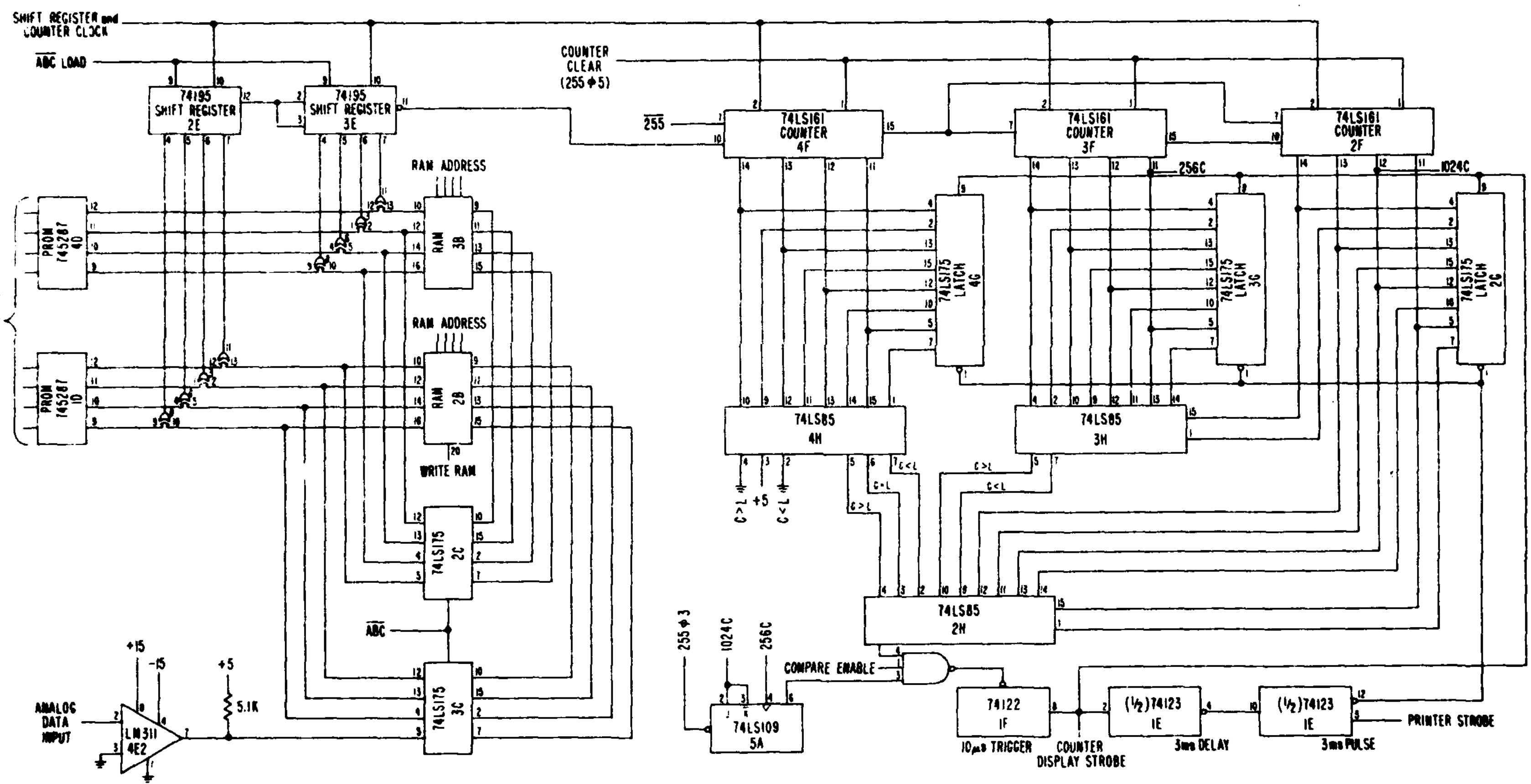


Figure A3. Correlation detector.

APPENDIX B: TEST CONFIGURATIONS

The 8 test conditions outlined in Table 1 are briefly discussed in this section.

(a) Filters

The filters used were 4 pole bandpass filters with the 6-dB bandwidths of 300 Hz - 3 kHz for the narrowband mode (N) and 200 Hz - 10 kHz for the wideband mode (W). The F1A weighing is discussed in ITT (1969).

(b) Pre-emphasis/De-emphasis

Figure B1 shows the results of several measurements of the 6-dB/octave pre-emphasis, the 6 dB de-emphasis, and the combined effect. Figure B2 shows only the pre-emphasis curve for the 3-dB/octave setting.

(c) Noise

The noise added to the signal was white gaussian noise from a broadband noise generator. The noise power in the received bandwidth was measured on an rms voltmeter, as was the signal without noise. An additional measurement of signal plus interference plus distortion to interference plus distortion, was made using a 1-kHz tone with the same rms voltage level as the speech, and using a notch filter with 60 dB attenuation on 1 kHz. Figure B3 shows a block diagram of this system.

(d) Clipping

The clipping was accomplished using the network diagrammed in Figure B4. The rms audio signal voltage was preamplified to 30 dB above the clipping level. The signal plus interference plus distortion to interference plus distortion ratio

$$\frac{S+I+D}{I+D}$$

is determined by the harmonics, and would be 7 dB if the clipping produced a square wave.

(e) Crosstalk

The intelligible crosstalk used was prerecorded speech taken from AM radio broadcasts. The same segment was used during each of the four test conditions and included three contiguous time frames; each with a different speaker, one at the average level, one 3 dB (approximately) below the average level and one 3 dB (approximately) above the average level for the entire segment.

(f) Sine Wave Interference

A 2-kHz sine wave was used as interference in test setup #8. The interference ratios are the ratios of the rms voltages at the input to the recorder.

(g) Recording

The voice signals were recorded on a 1/2 inch tape recorder at 15 ips. Both the clear signal and the noisy signal were recorded on separate tracks. These signals were then transcribed onto 1/4 inch tape (reel-to-reel) for the subjective scoring which was done by the U.S. Army Electronics Proving ground at Ft. Huachuca, AZ. The 1/2 inch tapes were used in the data analysis described in Section 4.

REFERENCE

ITT (1969), Reference Data for Radio Engineers, Fifth Edition, (Howard W. Sams and Co. Inc., New York, NY).

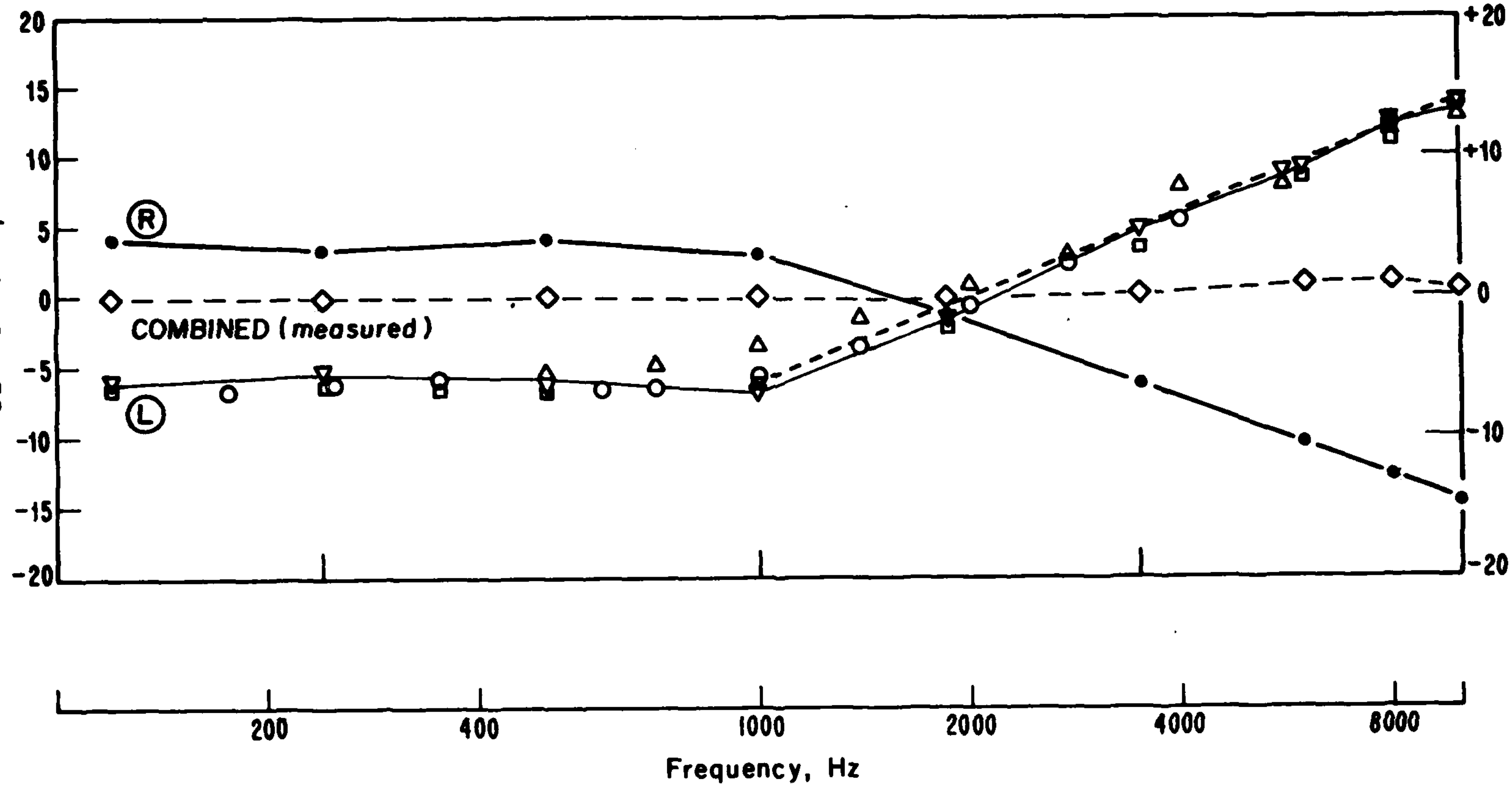


Figure B1. Measurement of the pre-emphasis (L), de-emphasis (R), and combined (dashed) characteristics for the 6-dB/octave above 1-Hz case. (The different symbols on the (L) curve are measurements made on different days.)

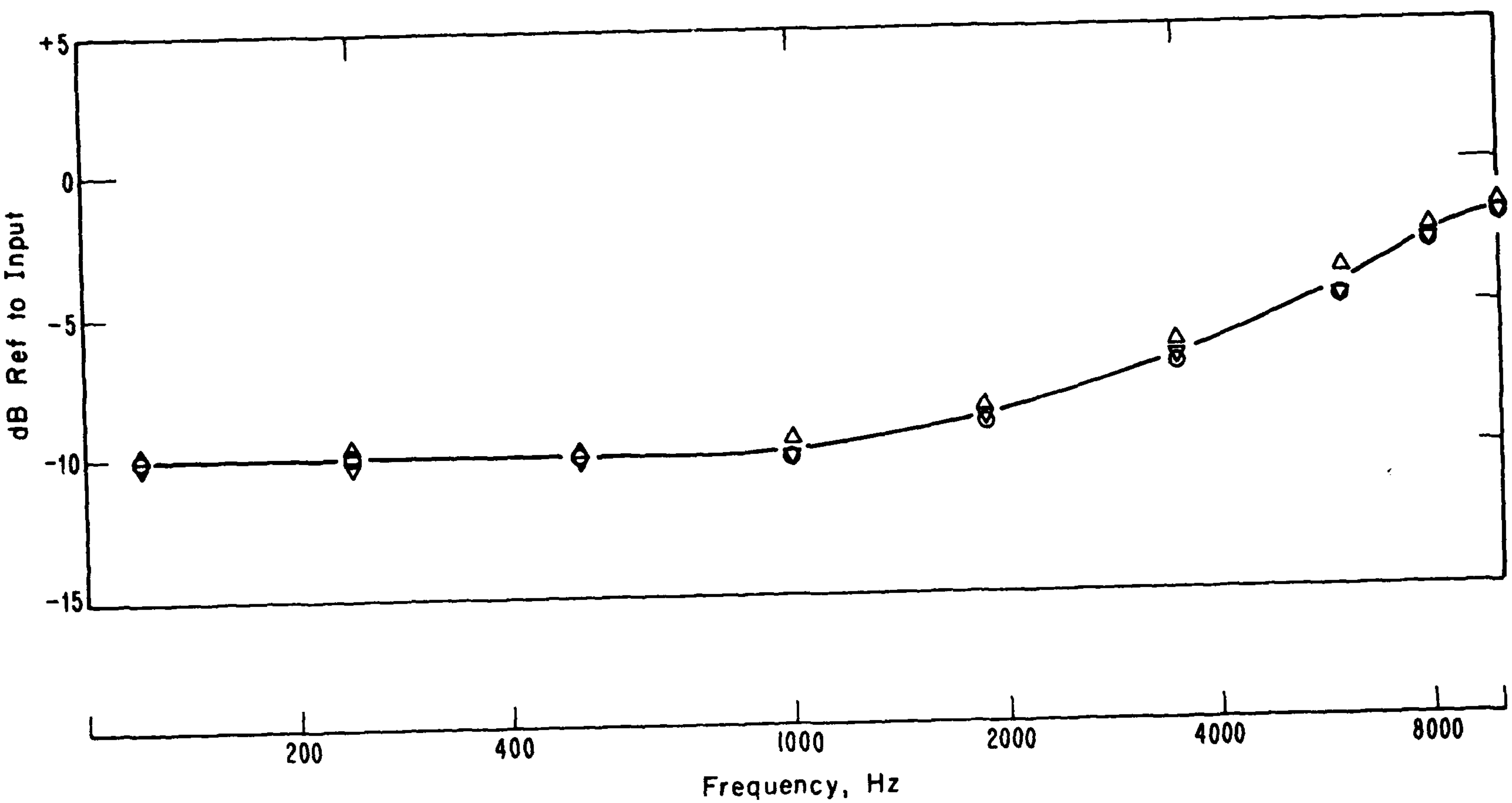


Figure B2. Measurement of the 3-dB/octave pre-emphasis characteristics.

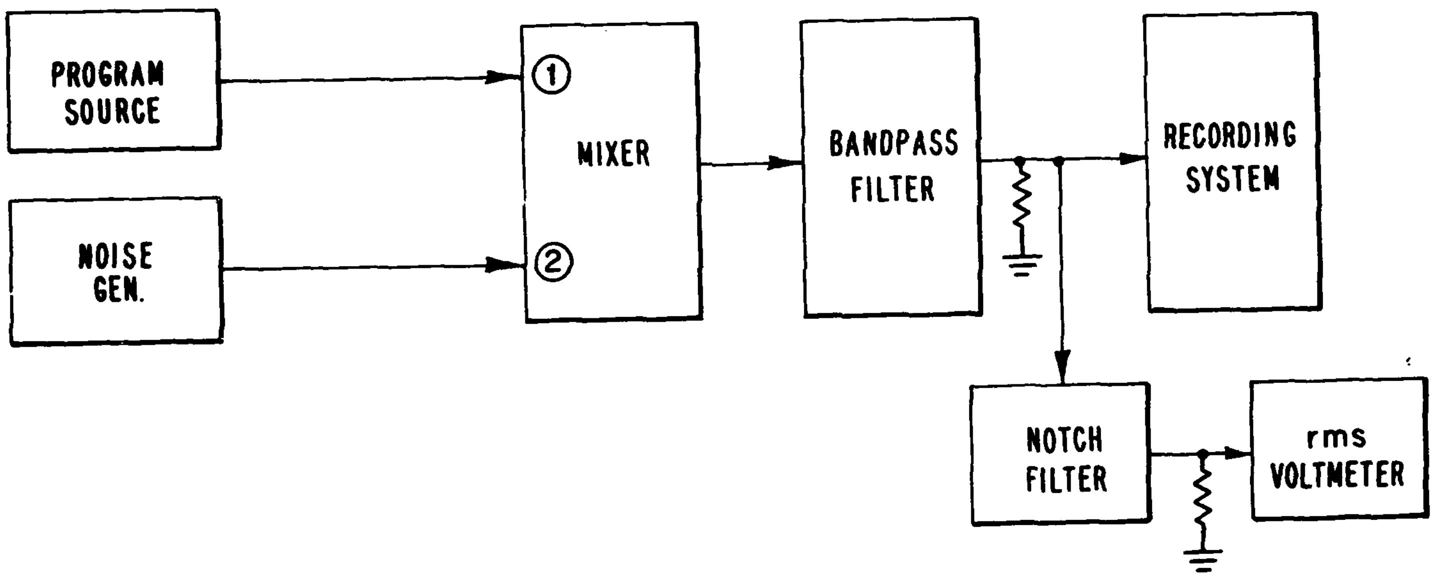


Figure B3. Block diagram for test #1. The notch filter and rms voltmeter are used for the measurement of $\frac{S+I+D}{I+D}$.

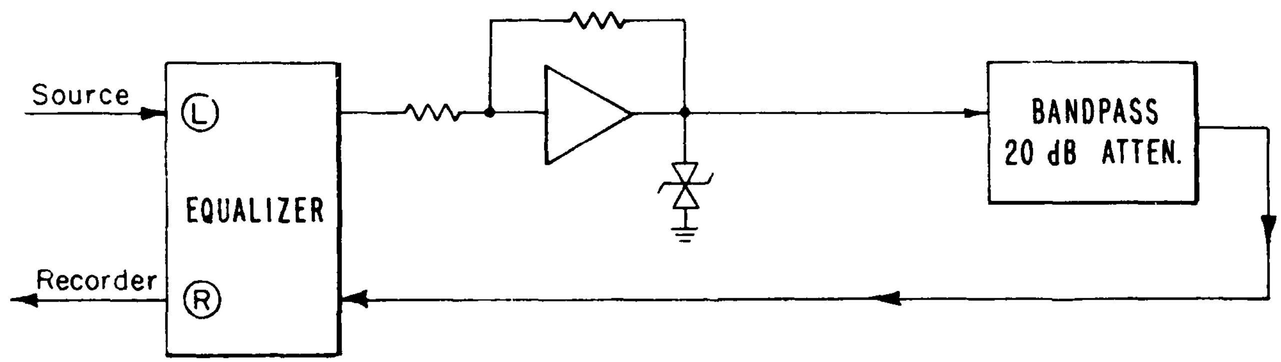


Figure B4. Block diagram for the hard limiter.